

Use of non-homogeneous Poisson process (NHPP) in presence of change-points to analyze drought periods: a case study in Brazil

Jorge Alberto Achcar¹ · Emílio Augusto Coelho-Barros² · Roberto Molina de Souza²

Received: 10 September 2015 / Revised: 15 February 2016 / Published online: 21 March 2016
© Springer Science+Business Media New York 2016

Abstract Rain precipitation in the last years has been very atypical in different regions of the world, possibly, due to climate changes. We analyze Standard Precipitation Index (SPI) measures (1, 3, 6 and 12-month timescales) for a large city in Brazil: Campinas located in the southeast region of Brazil, São Paulo State, ranging from January 01, 1947 to May 01, 2011. A Bayesian analysis of non-homogeneous Poisson processes in presence or not of change-points is developed using Markov Chain Monte Carlo methods in the data analysis. We consider a special class of models: the power law process. We also discuss some discrimination methods for the choice of the better model to be used for the rain precipitation data.

Keywords Bayesian inference · Change-points · Non-homogeneous Poisson processes · Power law process · Rain precipitation · Standard Precipitation Index

Handling Editor: Bryan F. J. Manly.

✉ Emílio Augusto Coelho-Barros
eabarros@utfpr.edu.br

Jorge Alberto Achcar
achcar@fmrp.usp.br

Roberto Molina de Souza
rmolinasouza@utfpr.edu.br

¹ Departamento de Medicina Social, FMRP, Universidade de São Paulo, Ribeirão Preto, SP, Brazil

² Departamento de Matemática, Universidade Tecnológica Federal do Paraná, Cornélio Procopio, PR, Brazil

1 Introduction

In recent years it has been observed that rain precipitation had an atypical behavior, often with excessive rainfall in some areas and very little rain in other regions, perhaps caused by climate changes observed on the planet. In Brazil this atypical behavior has been observed in the southeast region where it was rarely seen many drought events and short of water in water reservoirs that supply the big cities and agriculture. This temporal behavior of large drought events is common in northeastern region of Brazil, but rarely seen in the southeast region, the region with the highest population, highest industrialization and significant agricultural production.

A drought event results from lower levels of precipitations compared to standard normal precipitations observed during a long period of time. To monitor drought events, the literature present different methodologies such as percentage of normal precipitation and precipitation percentiles or the Palmer Drought Severity Index (PDSI), that is a measurement of dryness based on recent precipitation and temperature (Palmer 1965).

A popular index introduced in the literature and used by many countries around the world, more simple, easy to calculate, statistically relevant and effective in analyzing wet periods/cycles or dry periods/cycles is the Standardized Precipitation Index (SPI) introduced by McKee et al. (1993, 1995). This index has been extensively used by climatologists around the world (Karavitis 1998, 1999; Hayes et al. 1999; Lana et al. 2001; Sönmez et al. 2005; Livada and Assimakopoulos 2007; Chortaria et al. 2010; Karavitis et al. 2011).

To calculate the SPI, it is required at least 20–30 years of monthly values; optimal measures requires 50–60 years (Guttman 1994). The SPI is based on the probability of precipitation for any time scale. The probability of observed precipitation is then transformed into an index.

The SPI quantifies the precipitation deficit for different timescales. These timescales reflect the impact of a drought event on the availability of the different water resources. Soil moisture conditions respond to precipitation anomalies on a relatively short scale; groundwater, streamflow and reservoir storage reflect the longer-term precipitation anomalies. For these reasons, McKee et al. (1993) originally calculated the SPI for 3, 6, 12, 24 and 48-month timescales.

The SPI calculation for any location is based on the long-term precipitation record for a specified period. The SPI index is based on the cumulative probability of a rainfall occurring at a climate station. The historic rainfall data of the station is fitted to an asymmetrical distribution as a gamma distribution (usually this distribution gives good fit for the rainfall data) with the gamma distribution parameters being estimated by maximum likelihood estimation method. In this way, based on the historic rainfall data, an analyst can find the probability of the rainfall being less than or equal to a certain amount. Thus, the probability of rainfall being less than or equal to the median rainfall for a specified area will be about 0.5, while the probability of rainfall being less than or equal to an amount much smaller than the median will be lower. Therefore if a particular rainfall event gives a low probability on the cumulative probability function, then this is indicative of a likely drought event. Alternatively, a rainfall event which gives a high probability on the cumulative probability function is an anomalously wet event. In this

Table 1 SPI values

2.0+	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
−0.99 to 0.99	Near normal
−1.0 to −1.49	Moderately dry
−1.5 to −1.99	Severely dry
−2 and less	Extremely dry

way the distribution have a standard deviation and a mean which depends on the rainfall characteristics of a specified area. Therefore it is difficult to compare rainfall events for two or more different areas in terms of drought, as drought is really a “below-normal” rainfall event. It is important to point out that what is “normal rainfall” in one area can be surplus rainfall in another area, taking in account only rainfall amounts. The SPI index is obtained from a transformation of the obtained cumulative probability gamma function into a standard normal random variable Z with mean of zero and standard deviation of one. A new variate is formed, and the transformation is done in such a way that each rainfall amount in the old (gamma) function has a corresponding value in the new (transformed) Z function. The probability that the rainfall is less than or equal to any rainfall amount will be the same as the probability that the new variate is less than or equal to the corresponding value of that rainfall amount. The resultant transformed variate is always in the same units (Edwards 1997). Positive SPI values indicate greater than median precipitation and negative values indicate less than median precipitation. Because the SPI is normalized, wetter and drier climates can be represented in the same way; thus, wet periods can also be monitored using the SPI.

A classification system introduced by McKee et al. (1993) (see Table 1) is used to define drought event intensities obtained from the SPI. They also defined the criteria for a drought event for different timescales. A drought event occurs any time the SPI is continuously negative and reaches an intensity of -1.0 or less. The event ends when the SPI becomes positive. Each drought event, therefore, has a duration defined by its beginning and end, and an intensity for each month that the event continues. The positive sum of the SPI for all the months within a drought event can be termed the drought’s “magnitude”.

In this paper, we consider as a main goal of the study, the SPI measures (1, 3, 6 and 12-month timescales) for a large industrial city in Brazil, Campinas, with more than 1 million people located in the southeast region of Brazil, São Paulo State, ranging from January 01, 1947 to May 01, 2011. The data set is available in the site of the CIIAGRO (Índice Padronizado de Precipitação—SPI), a site related to climate in São Paulo State, Brazil: <http://www.ciiagro.sp.gov.br/ciiagroonline/Listagens/SPI/LspiLocal.asp>.

In Fig. 1, we have the time series plots for the $SPI - 1$, $SPI - 3$, $SPI - 6$ and $SPI - 12$ in Campinas for the period ranging from January 01, 1947 to May 01, 2011.

From the plots of Fig. 1, we observe very atypical values of SPI (< -2.0) indicating severely dry period from the 600th to the 620th months which corresponds to the period ranging from December 1996 to August 1998. That is, we observe a great variability of the SPI measures in this period of 2 years with alarming consequences.

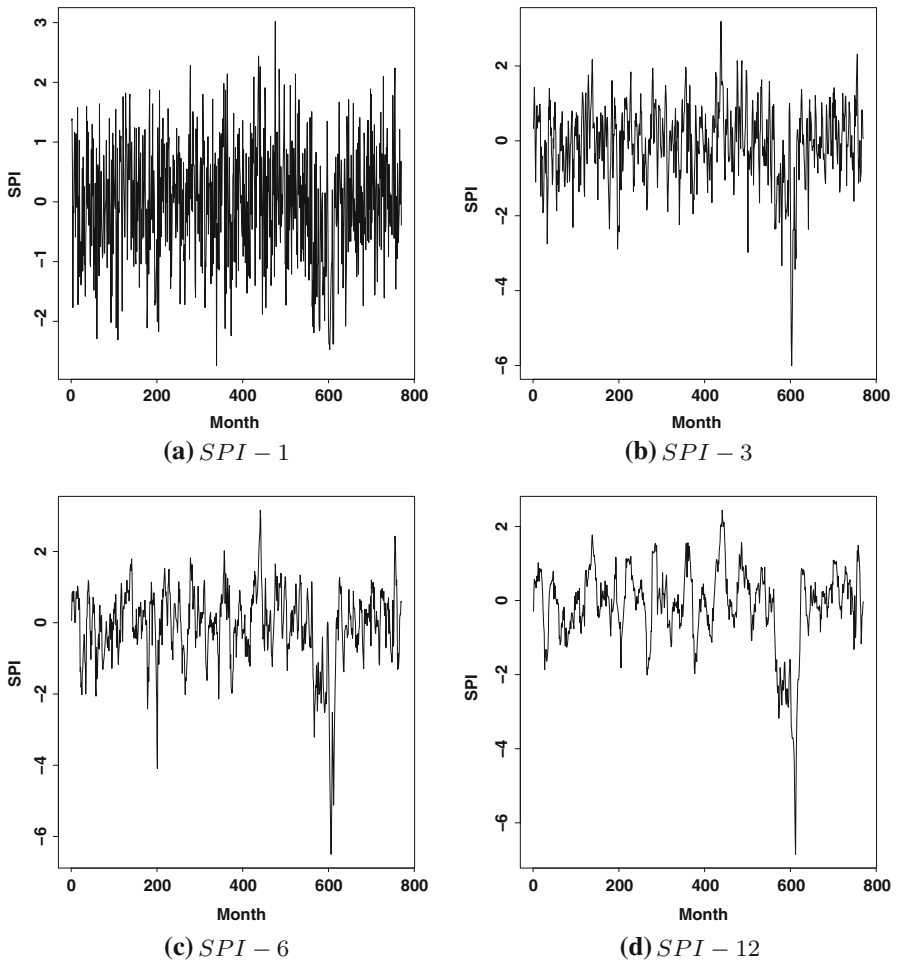


Fig. 1 Campinas SPI measures. **a** $SPI - 1$, **b** $SPI - 3$, **c** $SPI - 6$ and **d** $SPI - 12$

It is important to point out that the period from December 1996 to August 1998 is a period of a record breaking El Niño event (see NOAA technical report for the El Niño effects in North America in <http://www1.ncdc.noaa.gov/pub/data/techrpts/tr9802/tr9802>), which could be connected to this drought in Brazil.

To model this time series SPI data, we could consider a time series modeling approach as for example using stochastic volatility (SV) models (Ghysels 1996; Kim et al. 1998; Meyer and Yu 2000) with a Bayesian hierarchical approach. Alternatively, in place to model the SPI time series, since a drought event occurs any time the SPI is continuously negative and reaches an intensity of -1.0 or less, we consider the calendar time or epochs of droughts less or equal to the threshold $L = -1.0$ for Campinas, during the period ranging from January 01, 1947 to May 01, 2011, which corresponds to a period of $T = 770$ months.

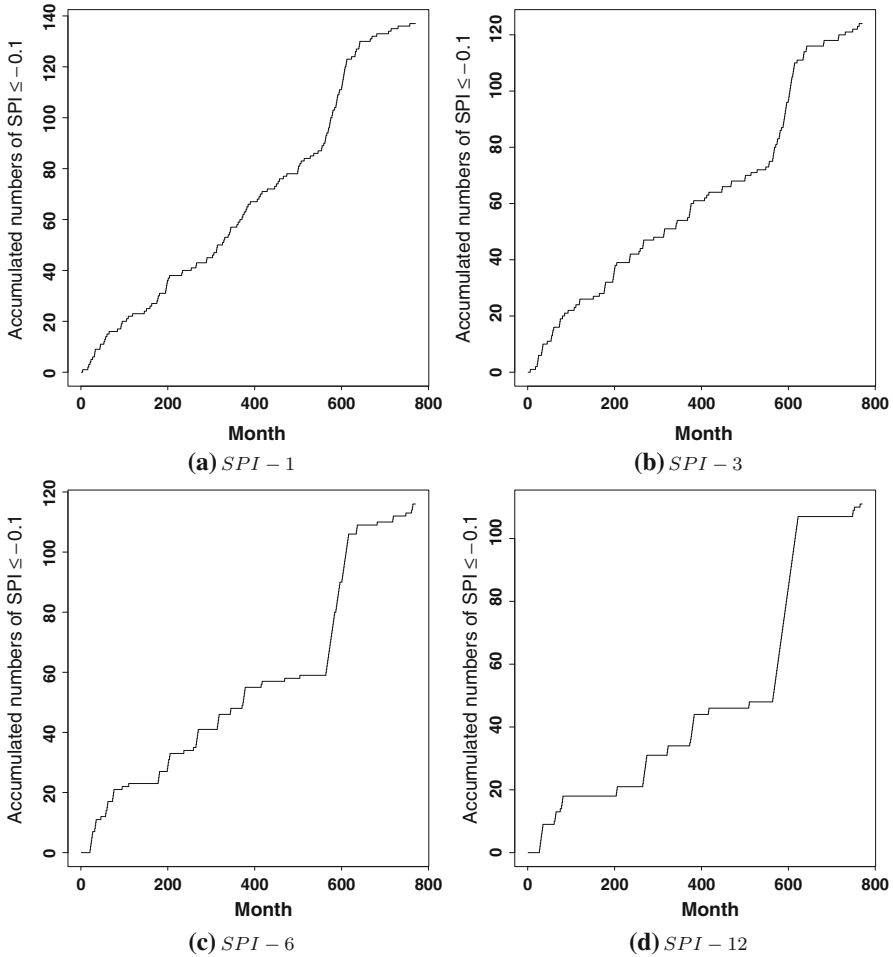


Fig. 2 Accumulated numbers of $SPI \leq -1.0$ in each time versus time (month of occurrence). **a** $SPI - 1$, **b** $SPI - 3$, **c** $SPI - 6$ and **d** $SPI - 12$

In Fig. 2, we have the plots of the accumulated numbers of drought event occurrences for $SPI - 1$, $SPI - 3$, $SPI - 6$ and $SPI - 12$ series ($SPI \leq -1.0$) in each time versus time (month of occurrence).

From Fig. 2, we observe the existence of one or two change-points close to the time period ranging from 600 to 620, which corresponds to approximately the period of time between the years 1996 and 1998.

For modeling the number of violation occurrences of drought events in Campinas city, we consider a point process to count violations. Let $N(t)$ be the cumulative number of violations that are observed during the interval $(0, T)$ and assume that $N(t)$ is modeled by a non-homogeneous Poisson process (NHPP). To have the intensity function $\lambda(t) = \frac{\partial m(t)}{\partial t} = \frac{\partial E[N(t)]}{\partial t}$ where $m(t)$ is the mean value function, a monotonic function of t , we could consider different parametrical forms introduced

in the literature (Cox and Lewis 1966; Musa and Okumoto 1984; Musa et al. 1987). Bayesian inference for non-homogeneous Poisson processes has been discussed by many authors (Kuo and Yang 1996; Pievatolo and Ruggeri 2004; Cid and Achcar 1999).

Alternatively, in many applications, we could have the presence of change-points for the intensity function of non-homogeneous Poisson processes. In this situation, we have interest to get inference on these change-points where the non-homogeneous Poisson process changes. Bayesian methods have been used by different authors to get inference for change-points models (Matthews and Farewell 1982; Achcar and Bolfarine 1989; Carlin et al. 1992; Dey and Purkayastha 1997; Achcar and Loibel 1998). Raftery and Akman (1986) considered a Bayesian analysis for Homogeneous Poisson Processes (HPP) in the presence of a change-point. Ruggeri and Sivaganesan (2005) introduces a Bayesian analysis for change-points in non-homogeneous Poisson processes considering Power Law Processes (PLP). Applications of non-homogeneous Poisson processes in presence or not of change-points are introduced by Achcar et al. (2008, 2010) to analyse ozone pollution data.

In this paper, we consider the use of Markov Chain Monte Carlo (MCMC) methods (Gelfand and Smith 1990; Smith and Roberts 1993; Chib and Greenberg 1995) to develop a Bayesian analysis for the change-point in NHPP assuming a popular class of models commonly used in software reliability or operating safety policy: the PLP processes.

The paper is organized in the following way: in Sect. 2, we introduce the likelihood function in presence or not of change-points; in Sect. 3, we present a Bayesian formulation of the model; in Sect. 4, we present a Bayesian analysis of the SPI data from Campinas city; finally in Sect. 5, we present some discussion of the obtained results.

2 The likelihood function

Let $\{N_t\}$, $t \geq 0$, be a non-homogeneous Poisson process with mean value function $m(t; \theta)$, where θ is a vector of parameters, representing the expected number of events (Barlow and Hunter 1960) (here is the number of drought months to time t). The full characterization of a process of this type is achieved by specification of the functional form of $m(t; \theta)$, or equivalently, of its intensity function $\lambda(t) = \frac{\partial m(t; \theta)}{\partial t}$.

We consider the use of a special case of NHPP's: the power law process (PLP) defined (Crow 1974), by the mean value function,

$$m_{PLP}(t; \theta) = \left(\frac{t}{\sigma}\right)^\alpha, \quad \alpha, \sigma > 0, \quad (1)$$

where $\theta = (\alpha, \sigma)$. The intensity function of this processes is given by,

$$\lambda_{PLP}(t; \theta) = \left(\frac{\alpha}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1}. \quad (2)$$

From 2, we observe that the intensity function $\lambda_{PLP}(t; \theta)$ has a flexible behavior for the PLP according to the value of α . The intensity function is increasing for $\alpha > 1$, decreasing for $\alpha < 1$ and constant (i.e. the NHPP is actually a Homogeneous Poisson Process) for $\alpha = 1$.

This process is considered to model the epochs of occurrence of drought violations up to time T . The data set is denoted by $D_T = \{n; t_1, \dots, t_n; T\}$ where n is the number of observed occurrence time t_i , which we suppose to be ordered, that is, $0 < t_1 < \dots < t_n < T$. Letting $N(t)$ denote the observable number of violation occurrences in $(0, t]$, considered to be modeled as an NHPP, $N(s + t) - N(s)$ given θ has a Poisson distribution $P[m(s + t; \theta) - m(s; \theta)]$ for $t > 0$ and independent increments.

The likelihood function for θ considering the time truncated model is given by (Cox and Lewis 1966),

$$L(\theta; D_T) = \left[\prod_{i=1}^n \lambda(t_i) \right] \exp[-m(T)]. \tag{3}$$

When the counting process undergoes changes over the time range $(0, T)$, we have a single change-point τ making a transition between two NHPP’s processes. In this way, the intensity function of the overall process is given by,

$$\lambda(t; \theta) = \begin{cases} \lambda_1(t), & 0 \leq t \leq \tau \\ \lambda_2(t), & t > \tau \end{cases}. \tag{4}$$

where $\lambda_j(t) = \lambda(t; \theta_j)$, $j = 1, 2$ is related with the intensity functions defined in (2) and $\theta = (\alpha_1, \sigma_1, \alpha_2, \sigma_2, \tau)$. Equivalently, letting $m_j(t) = m(t; \theta_j)$, the corresponding mean value function is given by,

$$m(t; \theta) = \begin{cases} m_1(t), & 0 \leq t \leq \tau \\ m_1(\tau) + m_2(t) - m_2(\tau), & t > \tau \end{cases}. \tag{5}$$

Observe that considering PLP models in the presence of a change-point, the intensity function (4) is given by,

$$\lambda(t; \theta) = \begin{cases} \left(\frac{\alpha_1}{\sigma_1}\right) \left(\frac{t}{\sigma_1}\right)^{\alpha_1-1}, & 0 \leq t \leq \tau \\ \left(\frac{\alpha_2}{\sigma_2}\right) \left(\frac{t}{\sigma_2}\right)^{\alpha_2-1}, & t > \tau \end{cases}. \tag{6}$$

with corresponding mean value function (5) given by,

$$m(t; \theta) = \begin{cases} \left(\frac{t}{\sigma_1}\right)^{\alpha_1}, & 0 \leq t \leq \tau \\ \left(\frac{\tau}{\sigma_1}\right)^{\alpha_1} + \left(\frac{t}{\sigma_2}\right)^{\alpha_2} - \left(\frac{\tau}{\sigma_2}\right)^{\alpha_2}, & t > \tau \end{cases}. \tag{7}$$

Considering the data $D_T = \{n; t_1, \dots, t_{N(\tau)}, t_{N(\tau)+1}, \dots, t_n; T\}$ for model (4)–(5), the likelihood function for θ is given by,

$$L(\theta; D_T) = \left[\prod_{i=1}^{N(\tau)} \lambda_1(t_i) \right] \exp[-m_1(\tau)] \left[\prod_{i=N(\tau)+1}^{N(T)} \lambda_2(t_i) \right] \times \exp[-m_2(T) + m_2(\tau)], \tag{8}$$

where $N(s)$ stands for the number of failures in the interval $(0, s]$. Observe that (3) is a special case of (8) when there is not a change-point (e.g., $\tau = T$).

In presence of two change-points, τ_1 and τ_2 making a transition between three NHPP’s processes, the intensity function of the overall process is given by,

$$\lambda(t; \theta) = \begin{cases} \left(\frac{\alpha_1}{\sigma_1}\right) \left(\frac{t}{\sigma_1}\right)^{\alpha_1-1}, & 0 \leq t \leq \tau_1 \\ \left(\frac{\alpha_2}{\sigma_2}\right) \left(\frac{t}{\sigma_2}\right)^{\alpha_2-1}, & \tau_1 \leq t \leq \tau_2 \\ \left(\frac{\alpha_3}{\sigma_3}\right) \left(\frac{t}{\sigma_3}\right)^{\alpha_3-1}, & t > \tau_2 \end{cases} \tag{9}$$

where $\theta = (\alpha_1, \sigma_1, \alpha_2, \sigma_2, \alpha_3, \sigma_3, \tau_1, \tau_2)$, with corresponding mean value function given by,

$$m(t; \theta) = \begin{cases} \left(\frac{t}{\sigma_1}\right)^{\alpha_1}, & 0 \leq t \leq \tau_1 \\ \left(\frac{\tau_1}{\sigma_1}\right)^{\alpha_1} + \left(\frac{t}{\sigma_2}\right)^{\alpha_2} - \left(\frac{\tau_1}{\sigma_2}\right)^{\alpha_2}, & \tau_1 \leq t \leq \tau_2 \\ \left(\frac{\tau_1}{\sigma_1}\right)^{\alpha_1} + \left(\frac{t}{\sigma_3}\right)^{\alpha_3} - \left(\frac{\tau_2}{\sigma_3}\right)^{\alpha_3} + \left(\frac{\tau_2}{\sigma_2}\right)^{\alpha_2} - \left(\frac{\tau_1}{\sigma_2}\right)^{\alpha_2}, & t > \tau_2 \end{cases} \tag{10}$$

Assuming that the data is observed up to a total time T , where the epochs of occurrence of cases are denoted by $t_i, i = 1, \dots, n, 0 < t_1 < t_2 < \dots < t_{N(\tau_1)} < t_{N(\tau_1)+1} < \dots < t_{N(\tau_2)} < t_{N(\tau_2)+1} < \dots < t_n < T$, the likelihood function for θ in the presence of two change-points τ_1 and τ_2 is given by,

$$L(\theta) = \left[\prod_{i=1}^{N(\tau_1)} \lambda_1(t_i) \right] e^{-m_1(\tau_1)} \left[\prod_{i=N(\tau_1)+1}^{N(\tau_2)} \lambda_2(t_i) \right] \times e^{-m_2(\tau_2)+m_2(\tau_1)} \left[\prod_{i=N(\tau_2)+1}^{N(T)} \lambda_3(t_i) \right] e^{-m_3(T)+m_3(\tau_2)}. \tag{11}$$

3 A Bayesian analysis

In this section, we introduce a Bayesian analysis for the PLP model. Posterior summaries of interest are obtained using Markov Chain Monte Carlo (MCMC) methods. For a Bayesian analysis of the proposed model, we first assume NHPP’s not in the presence of a change-point. We assume a uniform distribution $U [a, b]$ in the interval $[a = 0, b = 100]$ for the parameters α and σ to have approximately non-informative priors. We further assume prior independence among the parameters.

The joint posterior distribution for θ given the data D_T is,

$$p(\theta | D_T) \propto p(\theta) L(\theta | D_T) \tag{12}$$

where $p(\theta)$ denotes the joint distribution and $L(\theta | D_T)$ is the likelihood function given in (3). Simulated samples for the joint posterior distribution for θ are obtained using standard MCMC methods, that is, from the full conditional posterior distributions $p(\theta_i | \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n, D_T)$ for $i = 1, \dots, n$ (Gelfand and Smith 1990). A great simplification is obtained using the library `R2jags` (Su and Yajima 2015) in software R (R Core Team 2015), where we only need to specify the joint distribution for the data and the prior distributions for the parameters.

In a second stage for the Bayesian analysis, we assume the PLP model in the presence of a change-point τ . In this case, we assume an uniform prior distribution for the change-point in the interval $(0, T)$, and uniform non-informative prior distributions for the parameters $\alpha_1, \alpha_2, \sigma_1$ and σ_2 , that is, $U [0, 100]$. We further assume prior independence among the parameters.

In presence of two change points, we assume uniform non-informative priors for the change points parameters $\tau_1 \sim U [400, 600]$ and $\tau_2 \sim U [600, T]$, and uniform non-informative priors $U [0, 100]$ for the parameters $\alpha_1, \alpha_2, \alpha_3, \sigma_1, \sigma_2$ and σ_3 .

4 Analysis of the SPI data from Campinas

Consider the precipitation data of Campinas corresponding to the drought occurrences for $SPI - 1, SPI - 3, SPI - 6$ and $SPI - 12$ ($SPI \leq -1.0$) from January 01, 1947 to May 01, 2011. This corresponds to a total of $T = 770$ months. The Bayesian analysis is made considering three cases. In the first case, we consider NHPP’s with PLP intensity function without the presence of one change-point. In the second case, we consider the same NHPP’s but now assuming the presence of a change-point. In the third case, we consider the same NHPP’s but now assuming the presence of two change-points.

The selection of the best model is made using some existing Bayesian adequacy measures such as the Deviance Information Criterion (DIC) (Spiegelhalter et al. 2002) which is an approximation for the Bayes factor. Smaller values of DIC indicate better models. We also discriminate the models by comparing plots of the accumulated numbers of $SPI \leq -1.0$ in each time with the estimated mean value functions versus time of occurrence. The Bayesian analysis for all models was made using the library `R2jags` (Su and Yajima 2015) in software R (R Core Team 2015). Convergence of the

Table 2 Posterior means (standard deviation) for α and σ (PLP–NHPP not presence of change points)

	$\hat{\alpha}$	$\hat{\sigma}$	DIC
<i>SPI</i> – 1	1.0526 (0.0877)	7.5730 (3.0005)	751.30
<i>SPI</i> – 3	1.0042 (0.0868)	6.7560 (2.7845)	705.25
<i>SPI</i> – 6	1.0498 (0.0939)	8.8409 (3.5996)	675.38
<i>SPI</i> – 12	1.2474 (0.1144)	18.3014 (6.2272)	652.47

Gibbs sampling algorithm was monitored by usual time series plots for the simulated samples and also using some existing Bayesian convergence methods considering different initial values (Gelman and Rubin 1992).

4.1 NHPP model not in the presence of change-points

Assuming the PLP model, we consider uniform non-informative prior distributions for the parameters α and σ considering the four SPI series (*SPI* – 1 with $n = 137$ observations, *SPI* – 3 with $n = 124$ observations, *SPI* – 6 with $n = 116$ observations and *SPI* – 12 with $n = 111$ observations). A burn-in sample of size 10,000 was considered to eliminate the effect of the initial values. The posterior summaries of interest and the Monte Carlo estimate for DIC based on 2000 simulated Gibbs samples (by taking every 100th simulated value) are given in Table 2.

4.2 NHPP model in the presence of one change-point

Assuming the PLP model in presence of one change-point, we also consider uniform non-informative prior distributions for the parameters $\alpha_1, \alpha_2, \sigma_1, \sigma_2$ and τ . A burn-in sample of size 10,000 was considered to eliminate the effect of the initial values. The posterior summaries of interest and the Monte Carlo estimate for DIC based on 2000 simulated Gibbs samples (by taking every 100th simulated value) are given in Table 3.

4.3 NHPP model in the presence of two change-points

Assuming the PLP model in presence of two change-points, we also consider uniform non-informative prior distributions for the parameters $\alpha_1, \alpha_2, \alpha_3, \sigma_1, \sigma_2, \sigma_3, \tau_1$ and τ_2 . A burn-in sample of size 10,000 was considered to eliminate the effect of the initial values. The posterior summaries of interest and the Monte Carlo estimate for DIC based on 2000 simulated Gibbs samples (by taking every 100th simulated value) are given in Table 4.

From the obtained Monte Carlo estimates for the Deviance Information Criterion (DIC) given in Tables 2, 3 and 4, we observe smaller values of DIC assuming the PLP

Table 3 Posterior means (standard deviation) for $\alpha_1, \alpha_2, \sigma_1, \sigma_2$ and τ (PLP–NHPP presence of one change point)

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\tau}$	DIC
<i>SPI</i> – 1	1.1984 (0.1035)	1.2936 (0.2662)	11.6938 (4.0724)	56.4024 (27.8015)	652.52 (16.9473)	739.52
<i>SPI</i> – 3	1.0803 (0.2432)	1.3638 (0.3056)	8.5861 (5.8506)	47.7410 (30.3880)	565.89 (167.57)	705.11
<i>SPI</i> – 6	0.7811 (0.0913)	1.1760 (0.3112)	3.5134 (2.1747)	13.9759 (16.1642)	558.49 (6.7836)	655.92
<i>SPI</i> – 12	0.8579 (0.1136)	1.0061 (0.2483)	7.0486 (4.1647)	6.1740 (8.2344)	559.93 (4.8884)	620.91

model in presence of two change-points, that is an indication of better fitted model for the SPI data set. We also observe better fit of the PLP model in presence of two change-points by comparing plots of the accumulated number of $SPI \leq -1.0$ in each time with the estimated mean value functions versus time of occurrence (see Fig. 3).

5 Discussion of the results

The use of non-homogeneous Poisson processes assuming a specified parametrical form for the intensity function could be very useful to analyze precipitation data of cities throughout the world and explain some atypical drought event periods. This data would provide information about the epochs of occurrence of drought event violations of environmental standards during a specific period of time, as is the case of the SPI measures (1, 3, 6 and 12-month timescales) for the Campinas city, Brazil.

With the use of Markov Chain Monte Carlo methods, we obtain in a simple way the posterior distribution summaries of quantities of interest. The use of the library R2jags in software R gives a great simplification in the computational work to simulate the samples from the posterior distributions of interest.

A common problem with precipitation or other environmental data is the presence of one or more change-points, possible due to climate changes in the last years. In this case, we also observe that better NHPP models could be of great use. It is interesting to point out that other parametrical forms for the intensity function in the NHPP could be used in a similar way as it was considered using PLP intensity forms.

Important interpretations could be given from the obtained results. Considering the PLP–NHPP in presence of two change points (the best fitted model), we observe from the results of Table 4, that: (i) The credibility intervals for $\alpha_j, j = 1, 3$ in the four SPI measures (*SPI* – 1, *SPI* – 3, *SPI* – 6 and *SPI* – 12) cover the value 1, thus behaving similarly to a homogeneous PP, that is, a constant rate for a corresponding exponential distribution for the inter-arrival times for the SPI measure threshold $L = -1.0$ (a normal standard of rainfalls) whereas α_2 does not cover the value 1, with Bayesian estimates close to the value 2 for all SPI measures, an increasing intensity (2) which indicates an atypical behavior of rainfall for the period between the

Table 4 Posterior means (standard deviation) for $\alpha_1, \alpha_2, \alpha_3, \sigma_1, \sigma_2, \sigma_3, \tau_1$ and τ_2 (PLP–NHPP presence of two change points)

<i>SPI</i> – 1	$\hat{\alpha}_1$	$\hat{\sigma}_1$	$\hat{\tau}_1$	<i>DIC</i> = 717.58
	0.9864 (0.1019)	6.4222 (2.9832)	552.0760 (11.2752)	
	$\hat{\alpha}_2$	$\hat{\sigma}_2$	$\hat{\tau}_2$	
	2.0600 (0.4228)	56.1728 (27.7165)	632.6092 (14.6925)	
	$\hat{\alpha}_3$	$\hat{\sigma}_3$		
	1.3428 (0.2717)	55.2291 (27.8700)		
<i>SPI</i> – 3	$\hat{\alpha}_1$	$\hat{\sigma}_1$	$\hat{\tau}_1$	<i>DIC</i> = 660.11
	0.8502 (0.0924)	3.9576 (2.2045)	558.8111 (6.0918)	
	$\hat{\alpha}_2$	$\hat{\sigma}_2$	$\hat{\tau}_2$	
	2.1417 (0.4297)	57.1688 (27.4622)	620.6246 (10.0222)	
	$\hat{\alpha}_3$	$\hat{\sigma}_3$		
	1.4342 (0.2791)	56.0449 (27.8015)		
<i>SPI</i> – 6	$\hat{\alpha}_1$	$\hat{\sigma}_1$	$\hat{\tau}_1$	<i>DIC</i> = 572.89
	0.7765 (0.0952)	3.4736 (2.2488)	562.8702 (1.4886)	
	$\hat{\alpha}_2$	$\hat{\sigma}_2$	$\hat{\tau}_2$	
	2.2520 (0.4569)	55.9966 (27.6367)	617.0276 (1.3207)	
	$\hat{\alpha}_3$	$\hat{\sigma}_3$		
	1.3676 (0.2575)	57.6732 (27.4421)		
<i>SPI</i> – 12	$\hat{\alpha}_1$	$\hat{\sigma}_1$	$\hat{\tau}_1$	<i>DIC</i> = 499.81
	0.8607 (0.1155)	7.2052 (4.2379)	563.0705 (1.2856)	
	$\hat{\alpha}_2$	$\hat{\sigma}_2$	$\hat{\tau}_2$	
	2.1979 (0.4483)	49.5556 (27.3415)	622.9102 (1.0153)	
	$\hat{\alpha}_3$	$\hat{\sigma}_3$		
	1.0946 (0.2306)	58.9024 (27.4333)		

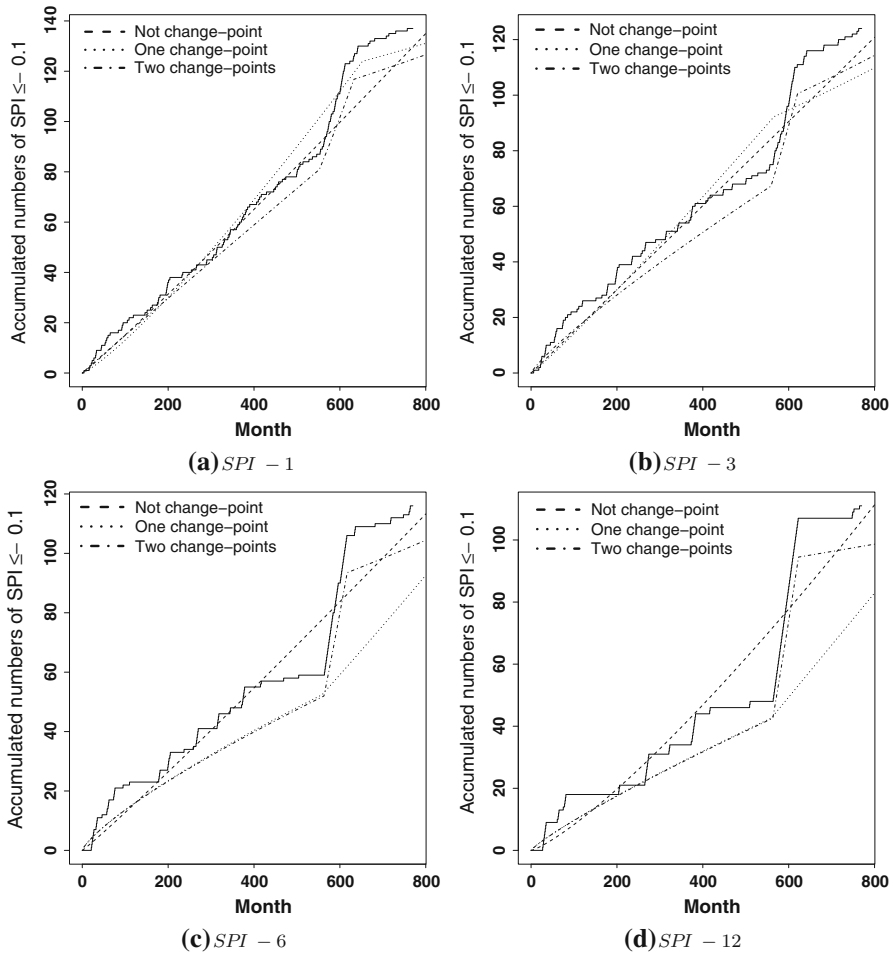


Fig. 3 Plots of the accumulated number of $SPI \leq -1.0$ in each time with the estimated mean value functions versus time of occurrence. **a** $SPI - 1$, **b** $SPI - 3$, **c** $SPI - 6$ and **d** $SPI - 12$

two change-points (larger period of droughts). This means that for the period ranging from the two estimated change-points given respectively by the approximations (four SPI measures) 560th and 620th months, corresponding to the period ranging from February, 1996 to August, 1998 we have many occurrences of the SPI measures less than the threshold $L = -1.0$ (similarly to a reliability deterioration model used in reliability studies). (ii) It is interesting to observe that the Bayesian estimate for σ_1 is close to 6.4 and that the Bayesian estimates for σ_2 and σ_3 are close to 56. Although it is hard to interpret this parameter, we could find estimates of the mean value function (10) for each specified value of time t .

It is important to point out that other modeling approaches could be considered for the climate dataset as the spline model introduced in Ramsay and Silverman’s (2006) Functional Data Analysis, but this modeling and comparison of results could

be the goal of a future paper. Finally, it is important to point out that the class of models used for the analysis of the rainfall dataset of Campinas could be used in many other applications with ecological or environmental data as air or water pollution data, temperature changes, level of sea changes and many other applications.

Acknowledgments The authors are very grateful to the Editor and a referee for their helpful and useful comments that improved the manuscript. The first author was partially supported by CNPq-Brazil.

References

- Achcar JA, Bolfarine H (1989) Constant hazard against a change-point alternative: a Bayesian approach with censored data. *Commun Stat Theory Methods* 18(10):3801–3819. doi:[10.1080/03610928908830124](https://doi.org/10.1080/03610928908830124)
- Achcar JA, Fernández-Bremauntz AA, Rodrigues ER, Tzintzun G (2008) Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model. *Environmetrics* 19(5):469–485. doi:[10.1002/env.890](https://doi.org/10.1002/env.890)
- Achcar JA, Loibel S (1998) Constant hazard models with a change-point: a Bayesian analysis using Markov chain Monte Carlo methods. *Biom J* 40:543–555
- Achcar JA, Rodrigues ER, Paulino CD, Soares P (2010) Non-homogeneous Poisson models with a change-point: an application to ozone peaks in Mexico City. *Environ Ecol Stat* 17(4):521–541. doi:[10.1007/s10651-009-0114-3](https://doi.org/10.1007/s10651-009-0114-3)
- Barlow R, Hunter L (1960) Optimum preventive maintenance policies. *Oper Res* 8:90–100
- Carlin BP, Gelfand AE, Smith AF (1992) Hierarchical Bayesian analysis of changepoint problems. *Appl Stat* 41(2):389–405
- Chib S, Greenberg E (1995) Understanding the Metropolis–Hastings algorithm. *Am Stat* 49(4):327–335
- Chortaria C, Karavitis CA, Alexandris S (2010) Development of the spi drought index for greece using geo-statistical methods. In: Proceedings of BALWOIS 2010 international conference
- Cid JER, Achcar JA (1999) Bayesian inference for nonhomogeneous Poisson processes in software reliability models assuming nonmonotonic intensity functions. *Comput Stat Data Anal* 32(2):147–159
- Cox DR, Lewis PAW (1966) The statistical analysis of series of events. Methuen & Co., Ltd.; Wiley, London; New York
- Crow LH (1974) Reliability analysis for complex systems. In: Proschan F, Serfling RJ (eds) Reliability and biometry. SIAM, Philadelphia, pp 379–410
- Dey DK, Purkayastha S (1997) Bayesian approach to change point problems. *Commun Stat Theory Methods* 26(8):2035–2047. doi:[10.1080/03610929708832030](https://doi.org/10.1080/03610929708832030)
- Edwards DC (1997) Characteristics of 20th century drought in the united states at multiple time scales. Tech. rep., DTIC Document
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85(410):398–409
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7(4):457–472
- Ghysels E (1996) Stochastic volatility. Statistical methods on finance. Springer, New York
- Guttman NB (1994) On the sensitivity of sample l moments to sample size. *J Clim* 7(6):1026–1029
- Hayes MJ, Svoboda MD, Wilhite DA, Vanyarkho OV (1999) Monitoring the 1996 drought using the standardized precipitation index. *Bull Am Meteorol Soc* 80(3):429–438
- Karavitis CA (1998) Drought and urban water supplies: the case of metropolitan athens. *Water Policy* 1(5):505–524
- Karavitis CA (1999) Decision support systems for drought management strategies in metropolitan athens. *Water Int* 24(1):10–21
- Karavitis CA, Alexandris S, Tsemelis DE, Athanasopoulos G (2011) Application of the standardized precipitation index (spi) in Greece. *Water* 3(3):787–805
- Kim S, Shephard N, Chib S (1998) Stochastic volatility: likelihood inference and comparison with arch models. *Rev Econ Stud* 65(3):361–393
- Kuo L, Yang TY (1996) Bayesian computation for nonhomogeneous Poisson processes in software reliability. *J Am Stat Assoc* 91(434):763–773

- Lana X, Serra C, Burgueño A (2001) Patterns of monthly rainfall shortage and excess in terms of the standardized precipitation index for catalonia (ne spain). *Int J Climatol* 21(13):1669–1691
- Livada I, Assimakopoulos V (2007) Spatial and temporal analysis of drought in Greece using the standardized precipitation index (spi). *Theor Appl Climatol* 89(3–4):143–153
- Matthews DE, Farewell VT (1982) On testing for a constant hazard against a change-point alternative. *Biometrics* 38(2):463–468
- McKee TB, Doesken NJ, Kleist J (1995) Drought monitoring with multiple time scales. In: Ninth conference on applied climatology. American Meteorological Society, Boston
- McKee TB, Doesken NJ, Kleist J et al (1993) The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th conference on applied climatology, vol 17. American Meteorological Society, Boston, MA, USA, pp 179–183
- Meyer R, Yu J (2000) Bugs for a Bayesian analysis of stochastic volatility models. *Econ J* 3(2):198–215
- Musa JD, Iannino A, Okumoto K (1987) Software reliability: measurement, prediction, application. McGraw-Hill Inc, New York
- Musa JD, Okumoto K (1984) A logarithmic poisson execution time model for software reliability measurement. In: Proceedings of the 7th international conference on software engineering, IEEE Press, pp 230–238
- Palmer WC (1965) Meteorological drought, weather bureau research paper no. 45. US Department of Commerce, Washington
- Pievatolo A, Ruggeri F (2004) Bayesian reliability analysis of complex repairable systems. *Appl Stoch Model Bus Ind* 20(3):253–264. doi:10.1002/asmb.522
- R Core Team(2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Raftery A, Akman V (1986) Bayesian analysis of a Poisson process with a change-point. *Biometrika* 71(1):85–89
- Ramsay JO, Silverman BW (2006) Functional data analysis. Wiley, New York
- Ruggeri F, Sivaganesan S (2005) On modeling change points in non-homogeneous Poisson processes. *Stat Inference Stoch Process* 8(3):311–329
- Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J R Stat Soc Ser B Methodol* 55(1): 3–23. [http://links.jstor.org/sici?sici=0035-9246\(1993\)55:1<3:BCVTGS>2.0.CO;2-#&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1993)55:1<3:BCVTGS>2.0.CO;2-#&origin=MSN)
- Sönmez FK, Koemuescue AU, Erkan A, Turgu E (2005) An analysis of spatial and temporal dimension of drought vulnerability in Turkey using the standardized precipitation index. *Nat Hazards* 35(2):243–264
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol* 64(4):583–639
- Su YS, Yajima M (2015) R2jags: Using R to Run 'JAGS'. R package version 0.5-6. <http://CRAN.R-project.org/package=R2jags>

Jorge Alberto Achcar is a Full Professor in Statistics of the Department of Social Medicine of the Medical School of the University of São Paulo (USP), Ribeirão Preto, SP, Brazil. He holds a Ph.D. in Statistics, Department of Statistics, University of Wisconsin, Madison, USA, 1982, a MSc in Statistics, Institute of Mathematics of the University of São Paulo, Brazil, 1976 and a BSc in Mathematics, University of São Paulo State (UNESP), Brazil, 1972.

Emílio Augusto Coelho-Barros is a Associate Professor in Statistics of the Department of Mathematics of the Federal Technological University of Paraná, Cornélio Procópio, PR, Brazil. First degree in Statistics from Universidade Estadual de Maringá, Brasil, 2005. Specialist in Biostatistics from the University of São Paulo, 2006. Ph.D. in Quantitative Methods in the area of Public Health from the Ribeirão Preto Medical School (USP), 2014.

Roberto Molina de Souza is a Associate Professor in Statistics of the Department of Mathematics of the Federal Technological University of Paraná, Cornélio Procópio, PR, Brazil. First degree in Statistics from Universidade Estadual de Maringá, Brasil, 2005. Specialist in Biostatistics from the University of São Paulo, 2006. Ph.D. in Quantitative Methods in the area of Public Health from the Ribeirão Preto Medical School (USP), 2015.